

On the optimality of a ℓ_1/ℓ_1 solver for sparse signal recovery from sparsely corrupted compressive measurements.

Laurent Jacques*

March 22, 2013

Abstract

This short note proves the $\ell_2 - \ell_1$ instance optimality of a ℓ_1/ℓ_1 solver, i.e., a variant of *basis pursuit denoising* with a ℓ_1 fidelity constraint, when applied to the estimation of sparse (or compressible) signals observed by sparsely corrupted compressive measurements. The approach simply combines two known results due to Y. Plan, R. Vershynin and E. Candès.

Conventions: Most of domain dimensions (e.g., M , N) are denoted by capital roman letters. Vectors and matrices are associated to bold symbols while lowercase light letters are associated to scalar values. The i^{th} component of a vector \mathbf{u} is u_i or $(\mathbf{u})_i$. The identity matrix is Id . The set of indices in \mathbb{R}^D is $[D] = \{1, \dots, D\}$. Scalar product between two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^D$ reads $\mathbf{u}^* \mathbf{v} = \langle \mathbf{u}, \mathbf{v} \rangle$ (using the transposition $(\cdot)^*$). For any $p \geq 1$, $\|\cdot\|_p$ represents the ℓ_p -norm such that $\|\mathbf{u}\|_p^p = \sum_i |u_i|^p$ with $\|\mathbf{u}\| = \|\mathbf{u}\|_2$ and $\|\mathbf{u}\|_\infty = \max_i |u_i|$. The ℓ_0 “norm” is $\|\mathbf{u}\|_0 = \#\text{supp } \mathbf{u}$, where $\#$ is the cardinality operator and $\text{supp } \mathbf{u} = \{i : u_i \neq 0\} \subseteq [D]$. For $\mathcal{S} \subseteq [D]$, $\mathbf{u}_{\mathcal{S}} \in \mathbb{R}^{\#\mathcal{S}}$ (or $\Phi_{\mathcal{S}}$) denotes the vector (resp. the matrix) obtained by retaining the components (resp. columns) of $\mathbf{u} \in \mathbb{R}^D$ (resp. $\Phi \in \mathbb{R}^{D' \times D}$) belonging to $\mathcal{S} \subseteq [D]$. The operator \mathcal{H}_K is the hard thresholding operator setting all the coefficients of a vector to 0 but those having the K strongest amplitudes. The set of canonical K -sparse signals in \mathbb{R}^N is $\Sigma_K = \{\mathbf{v} \in \mathbb{R}^N : \|\mathbf{v}\|_0 \leq K\}$. B_2^N and S^{N-1} are the ℓ_2 ball and $(N-1)$ -sphere in \mathbb{R}^N , respectively. Finally, the operator $\text{sign } \lambda$, which equals to 1 if λ is positive and -1 otherwise, is applied component wise onto vectors.

1 Introduction

Let us consider the case where a sparse (or compressible) signal $\mathbf{x} \in \mathbb{R}^N$ is observed with a random Gaussian matrix $\Phi \sim \mathcal{N}^{M \times N}(0, 1)$,

$$\mathbf{y} = \Phi \mathbf{x} + \mathbf{n}, \quad (1)$$

with a sparse (or Laplacian) noise \mathbf{n} of bounded ℓ_1 -power, i.e., there exists a bound $\epsilon > 0$ such that $\|\mathbf{n}\|_1 \leq \epsilon$ with high (and controlled) probability.

*L. J. is funded by the Belgian F.R.S-FNRS. ICTEAM Institute, ELEN Department, Université catholique de Louvain (UCL), B-1348 Louvain-la-Neuve, Belgium. Email: laurent.jacques@uclouvain.be

In this short note, we prove the stability of a variant of the *basis pursuit denoising* program, namely

$$\underset{\mathbf{u} \in \mathbb{R}^N}{\operatorname{argmin}} \|\mathbf{u}\|_1 \quad \text{s.t.} \quad \|\mathbf{y} - \Phi \mathbf{u}\|_1 \leq \epsilon, \quad (\text{BPDN-}\ell_1)$$

in estimating \mathbf{x} from \mathbf{y} under an ℓ_1 -fidelity constraint. The mathematical tools we are going to use are those developed in the recent work of Y. Plan and R. Vershynin in the context of 1-bit compressed sensing [1] combined with Candès' simplified proof of basis pursuit denoising $\ell_2 - \ell_1$ -instance optimality [2]. No elements are specially new except their combination. In particular, it is interesting to see how these two pieces of works fit nicely in order to reach the announced objective.

2 BPDN- ℓ_1 instance optimality

Here is the main result of this note.

Theorem 1. *Let $\Phi \in \mathbb{R}^{M \times N}$ be a sensing matrix used in (1) and assume that there exist 3 constants $\delta_{2K}, \delta_{3K} \in (0, 1)$ and $\nu > 0$ such that, for all $\mathbf{u} \in \Sigma_{2K}$ and $\mathbf{v} \in \Sigma_K$ with $\langle \mathbf{u}, \mathbf{v} \rangle = 0$,*

$$\left| \frac{1}{M} \|\Phi \mathbf{u}\|_1 - \nu \|\mathbf{u}\| \right| \leq \delta_{2K} \|\mathbf{u}\|, \quad (2)$$

$$\left| \frac{1}{M} \langle \operatorname{sign}(\Phi \mathbf{u}), \Phi \mathbf{v} \rangle \right| \leq \delta_{3K} \|\mathbf{v}\|. \quad (3)$$

Then, if $\delta_{2K} + \delta_{3K} \leq \nu - \frac{1}{2}$, the solution \mathbf{x}^ of BPDN- ℓ_1 respects*

$$\|\mathbf{x}^* - \mathbf{x}\| \leq 8 \frac{\epsilon}{M} + 12 e_0(K),$$

with $e_0(K) = \|\mathbf{x} - \mathbf{x}_K\|_1 / \sqrt{K}$.

Before to prove this theorem, the following lemma (mainly a rewriting of a result given in [1]) assures us on the feasibility of the conditions (2) and (3).

Lemma 1. *Let $N, M, K \in \mathbb{N}$ and $\delta \in [0, 1]$. There exist two constants $C, c > 0$ such that, for*

$$M \geq C \delta^{-6} K \log(2N/K) \quad (4)$$

and $\Phi \sim \mathcal{N}^{M \times N}(0, 1)$, we have, with a probability at least $1 - 8 \exp(-c\delta^2 M)$,

$$\left| \frac{1}{M} \|\Phi \mathbf{u}\|_1 - \sqrt{\frac{2}{\pi}} \|\mathbf{u}\| \right| \leq \delta \|\mathbf{u}\|, \quad (5)$$

$$\left| \frac{1}{M} \langle \operatorname{sign}(\Phi \mathbf{u}), \Phi \mathbf{v} \rangle \right| \leq \delta \|\mathbf{v}\|, \quad (6)$$

for all $\mathbf{u}, \mathbf{v} \in \Sigma_K$ with $\langle \mathbf{u}, \mathbf{v} \rangle = 0$.

Proof. Let us write $\mathcal{K} = \Sigma_K \cap B_2^N$ and $\mathcal{K}^* = \Sigma_K \cap S^{N-1}$. Using [1, Prop. 4.3] with $\tau = 0$, we know that there exist two constants $C, c > 0$ such that if

$$M \geq C \delta^{-6} K \log(2N/K)$$

and if $\Phi = (\varphi_1, \dots, \varphi_M)^T \sim \mathcal{N}^{M \times N}(0, 1)$ with $\varphi_i \in \mathbb{R}^N$ ($1 \leq i \leq M$), then, with probability at least $1 - 8 \exp(-c\delta^2 M)$,

$$\sup_{\mathbf{a} \in \mathcal{K}^*, \mathbf{b} \in \mathcal{K} - \mathcal{K}} |f_{\mathbf{a}}(\mathbf{b}) - \mathbb{E}f_{\mathbf{a}}(\mathbf{b})| \leq \delta,$$

where $f_{\mathbf{a}}(\mathbf{b}) := \frac{1}{M} \sum_j \text{sign}(\langle \varphi_j, \mathbf{a} \rangle) \langle \varphi_j, \mathbf{b} \rangle$. Knowing that $\mathbb{E}f_{\mathbf{a}}(\mathbf{b}) = \sqrt{\frac{2}{\pi}} \langle \mathbf{a}, \mathbf{b} \rangle$, this means that, under the same conditions,

$$\sup_{\mathbf{a} \in \mathcal{K}^*, \mathbf{b} \in \mathcal{K} - \mathcal{K}} \left| \frac{1}{M} \langle \text{sign}(\Phi \mathbf{a}), \Phi \mathbf{b} \rangle - \sqrt{\frac{2}{\pi}} \langle \mathbf{a}, \mathbf{b} \rangle \right| \leq \delta.$$

In particular, for any $\mathbf{u}, \mathbf{v} \in \Sigma_K$, since $\mathbf{u}/\|\mathbf{u}\| \in \mathcal{K}^*$ and $\mathbf{v}/\|\mathbf{v}\| \in \mathcal{K}^* \subset \mathcal{K} - \mathcal{K}$, we have

$$\left| \frac{1}{M} \langle \text{sign}(\Phi \mathbf{u}), \Phi \mathbf{v} \rangle - \sqrt{\frac{2}{\pi}} \|\mathbf{u}\|^{-1} \langle \mathbf{u}, \mathbf{v} \rangle \right| \leq \delta \|\mathbf{v}\|.$$

Therefore, if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$, $|\frac{1}{M} \langle \text{sign}(\Phi \mathbf{u}), \Phi \mathbf{v} \rangle| \leq \delta \|\mathbf{v}\|$, while taking $\mathbf{u} = \mathbf{v}$ leads to

$$\left| \frac{1}{M} \|\Phi \mathbf{u}\|_1 - \sqrt{\frac{2}{\pi}} \|\mathbf{u}\| \right| \leq \delta \|\mathbf{u}\|.$$

□

Remarks on δ : The dependency in δ^{-6} in (4) is probably not optimal and could be improved. This is actually due to the fact that this lemma is extendable to much more general sets than \mathcal{K} (e.g., compressible signals) [1]. For having only (5), [3, Lemma 5.3] shows that a dependency in δ^{-4} is allowed. Moreover, [4] shows that (5) holds of $M \geq M_0$ with $M_0 = O(\delta^{-2} K \log N/K)$. Proving that (6) is respected from the same number of measurements is an open problem.

Proof of Theorem 1. We follow partially the procedure given in [2] with an adaption due to the ℓ_1 -norm fidelity of BPDN- ℓ_1 . Let us write \mathbf{x}^* the solution of BPDN- ℓ_1 and $\mathbf{x}^* = \mathbf{x} + \mathbf{h}$. In order to bound the reconstruction error of BPDN- ℓ_1 , we have to characterize the behavior of $\|\mathbf{x}^* - \mathbf{x}\| = \|\mathbf{h}\|$.

We define $T_0 = \text{supp } \mathbf{x}_K$ and a partition $\{T_k : 1 \leq k \leq \lceil (N - K)/K \rceil\}$ of the support of $\mathbf{h}_{T_0^c}$. This partition is determined by ordering elements of \mathbf{h} off of the support of \mathbf{x}_K in decreasing absolute value. We have $|T_k| = K$ for all $k \geq 1$, $T_k \cap T_{k'} = \emptyset$ for $k \neq k'$, and crucially that $|h_j| \leq |h_i|$ for all $j \in T_{k+1}$ and $i \in T_k$.

We start from

$$\|\mathbf{h}\| \leq \|\mathbf{h}_{T_{01}}\| + \|\mathbf{h}_{T_{01}^c}\|, \quad (7)$$

with $T_{01} = T_0 \cup T_1$, and we are going to bound separately the two terms of the RHS. In [2], it is proved that

$$\|\mathbf{h}_{T_{01}^c}\| \leq \sum_{k \geq 2} \|\mathbf{h}_{T_k}\| \leq \|\mathbf{h}_{T_{01}}\| + 2e_0(K), \quad (8)$$

with $e_0(K) = \frac{1}{\sqrt{K}} \|\mathbf{x}_{T_0^c}\|_1$. Therefore,

$$\|\mathbf{h}\| \leq 2\|\mathbf{h}_{T_{01}}\| + 2e_0(K). \quad (9)$$

Let us bound now $\|\mathbf{h}_{T_{01}}\|$. We have

$$\|\Phi \mathbf{h}_{T_{01}}\|_1 = \langle \text{sign}(\Phi \mathbf{h}_{T_{01}}), \Phi \mathbf{h}_{T_{01}} \rangle = \langle \text{sign}(\Phi \mathbf{h}_{T_{01}}), \Phi \mathbf{h} \rangle - \sum_{k \geq 2} \langle \text{sign}(\Phi \mathbf{h}_{T_{01}}), \Phi \mathbf{h}_{T_k} \rangle.$$

By Hölder inequality,

$$\langle \text{sign}(\Phi \mathbf{h}_{T_{01}}), \Phi \mathbf{h} \rangle \leq \|\Phi \mathbf{h}\|_1 \leq \|\Phi \mathbf{x} - \mathbf{y}\|_1 + \|\Phi \mathbf{x} - \mathbf{y}\|_1 \leq 2\epsilon.$$

For any $k \geq 2$, since $\mathbf{h}_{T_{01}}$ and \mathbf{h}_{T_k} are $2K$ - and K -sparse, respectively, with $\langle \mathbf{h}_{T_{01}}, \mathbf{h}_{T_k} \rangle = 0$, we know from (3) that

$$|\langle \text{sign}(\Phi \mathbf{h}_{T_{01}}), \Phi \mathbf{h}_{T_k} \rangle| \leq M \delta_{3K} \|\mathbf{h}_{T_k}\|.$$

Therefore, using (2) and (8),

$$\begin{aligned} M(\nu - \delta_{2K}) \|\mathbf{h}_{T_{01}}\| &\leq \|\Phi \mathbf{h}_{T_{01}}\|_1 \leq 2\epsilon + M \delta_{3K} \sum_{k \geq 2} \|\mathbf{h}_{T_k}\| \\ &\leq 2\epsilon + M \delta_{3K} (\|\mathbf{h}_{T_{01}}\| + 2e_0(K)), \end{aligned}$$

or equivalently,

$$\|\mathbf{h}_{T_{01}}\| \leq \frac{2}{\nu - (\delta_{2K} + \delta_{3K})} \left(\frac{\epsilon}{M} + \delta_{3K} e_0(K) \right).$$

Using (9), we find,

$$\|\mathbf{h}\| \leq \frac{4}{\nu - (\delta_{2K} + \delta_{3K})} \frac{\epsilon}{M} + 4 \frac{\nu + \delta_{3K} - \delta_{2K}}{\nu - (\delta_{2K} + \delta_{3K})} e_0(K).$$

Finally, taking $\delta_{2K} + \delta_{3K} \leq \nu - \frac{1}{2}$ provides the result. \square

References

- [1] Y. Plan and R. Vershynin, “Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach,” *IEEE Transactions on Information Theory*, to appear, 2012.
- [2] E. Candès, “The restricted isometry property and its implications for compressed sensing,” *Compte Rendus de l’Academie des Sciences, Paris, Serie I*, vol. 346, pp. 589–592, 2008.
- [3] Y. Plan and R. Vershynin, “Dimension reduction by random hyperplane tessellations,” *arXiv preprint arXiv:1111.4452*, 2011.
- [4] L. Jacques, D. K. Hammond, and J. M Fadili, “Dequantizing compressed sensing: When oversampling and non-gaussian constraints combine,” *Information Theory, IEEE Transactions on*, vol. 57, no. 1, pp. 559–571, 2011.